Monitoring DUNE Data Streaming Efficiency

Lisa Paton, University of Alaska Anchorage

FERMILAB-POSTER-21-032-SCD-STUDENT

Introduction

DUNE computing
- 36 DUNE computing sites, 15 with storage
- Data streaming required for analysis and simulation

Purpose

Monitor/Analyze transfer data
- Who is running jobs
- Where is the data they're picking up
- How much data is being streamed
- How fast is the data being transferred

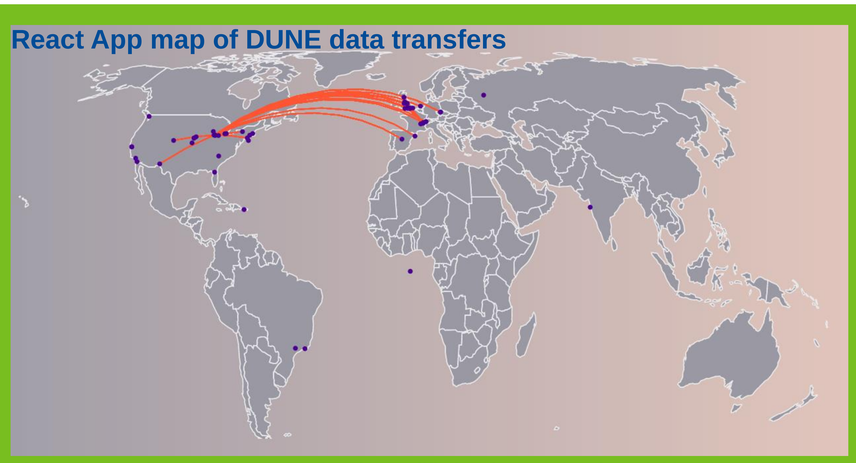
Methods and Materials

- React app
  - Tailored by Oregon State University students with Heidi Schellman
  - Kibana vs react
    - React will make network monitoring faster, more intuitive

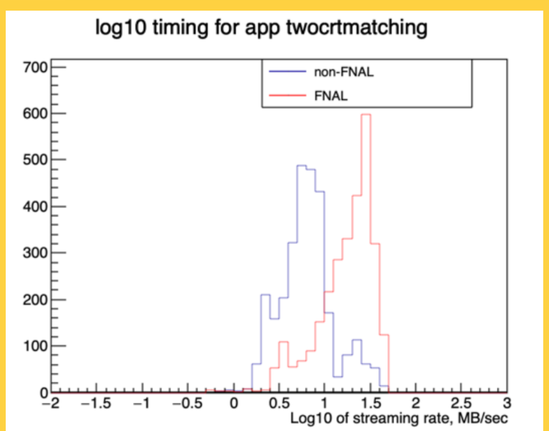
- SAM (sequential access metadata) & Rucio Data transfer systems
  - Kafka pushes SAM data into elastic search

- ElasticSearch
  - Scalable application that stores information with record-like structures for fast searching

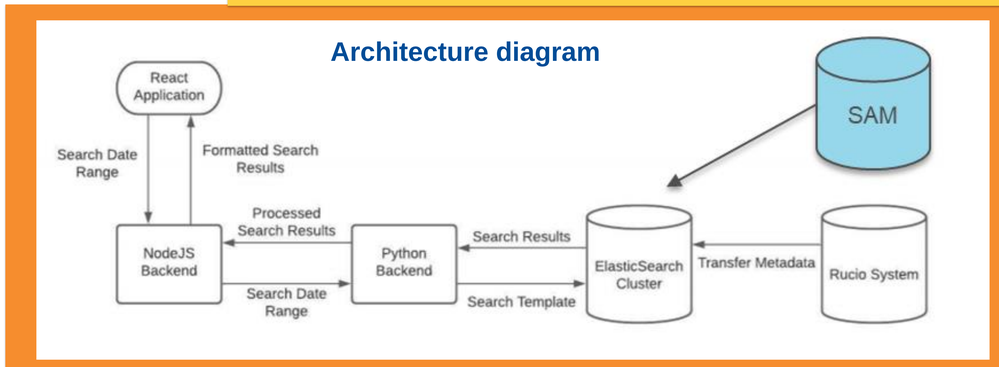
- Xrootd Streaming Data Collection
  - Query each SAM project from that day, separated by project id
    - Requires a new python backend
  - Comb through raw data files to make a summary of each day
  - Uses SAM states to determine duration, streaming rates, etc.
  - Make graphs and user csv file per day
  - Reformats user csv file to json to be fed into the React app
  - Concatenates daily user files
    - Speeds up react app by limited computing time from site's inputs



Above: Screenshot of React Map, using the modified Python Backend. Each blue dot represents a DUNE computing site, where the red lines connecting them display transferred data.



ROOT plot for streaming rates for the twocrtmatching (left) and reco processes, both on a log10 axis and separated between FNAL and non-FNAL (borrowed from Heidi Schellman's "Study of DUNE xrootd transactions")



This project makes a new python backend to query Xrootd transfers rather than Rucio transfers

Architecture diagram of the React application, flowing from right to left. The blue cylinder indicates the new addition of the SAM queries.

Results and Conclusions

ElasticSearch records, in combination with SAM project records can be mined to produce useful information about the characteristics of different DUNE applications. Within an application, the I/O rate can be reasonably well determined, both for local streaming and trans- Atlantic processing. For reconstruction, data location is reasonably unimportant as the jobs are CPU limited. For fast analysis, using the nearest data copy is important.

This manuscript has been authored by Fermi Research Alliance, LLC under Contract No. DE-AC02-07CH11359 with the U.S. Department of Energy, Office of Science, Office of High Energy Physics.